(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 May 2004 (27.05.2004)

PCT

(10) International Publication Number
WO 2004/044230 A1

(51) International Patent Classification[7]: C12Q 1/00

(21) International Application Number:
PCT/GB2003/004740

(22) International Filing Date:
4 November 2003 (04.11.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

0226441.4	12 November 2002 (12.11.2002)	GB
PCT/GB02/005112	12 November 2002 (12.11.2002)	GB
0308242.7	10 April 2003 (10.04.2003)	GB
0308238.5	10 April 2003 (10.04.2003)	GB

(71) Applicant *(for all designated States except US)*: **UNIVERSITY OF WALES COLLEGE OF MEDICINE** [GB/GB]; Heath Park, Cardiff CF4 4XN (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FRYKLUND, Linda** [SE/SE]; Pharmacia AB, Lindhagensgatan 100, S-112 87 Stockholm (SE). **LEWIS, Mark** [GB/GB]; University of Wales College of Medicine, Institute of Medical Genetics, Heath Park, Cardiff CF 14 4XN (GB). **COOPER, David** [GB/GB]; University of Wales College of Medicine, Institute of Medical Genetics, Heath Park, Cardiff CF14 4XN (GB).

(74) Agents: **FYLES, Julie, Marie** et al.; Wynne-Jones Laine & James, Morgan Arcade Chambers, 33 St Mary Street, Cardiff CF10 1AF (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: A METHODOLOGY OF ESTIMATING THE CONFORMATION OF A PROTEIN BY PROTEOLYSIS

(57) Abstract: The invention relates to a method for determining the effect of a polymorphism or mutation on the structural properties of a protein wherein the method relies on the structural properties of the protein and so its cleavage during proteolysis.

WO 2004/044230 A1

A METHODOLOGY OF ESTIMATING THE CONFORMATION OF A PROTEIN BY PROTEOLYSIS

The invention relates to a novel method for determining the significance of polymorphisms or mutations in a nucleic acid molecule encoding a protein.

Since the advent of gene sequencing technology in the late 1980's and the establishment of the human genome project in 1990 an enormous amount of information has been discovered about the sequence, or nature, of each gene in the human genome. Moreover, as the human genome project has developed the methods used to sequence genes have evolved considerably and this has led to the detection of variations within genes. Given that a typical gene could be 30 kilobases in length and that variations occur on average every 1100 bases, it follows that a tremendous amount of work needs to be undertaken in order to determine which variants are of clinical or technological significance. However, this is a prerequisite step if one is to exploit the knowledge available in the human genome project and so be in a position to understand, for example, the human condition, and particularly human diseases, and factors that may influence same and so lead to new therapies.

Typically, investigators in the field of human genetics who have obtained the sequence of the normal, or wild-type gene, set about looking for significant changes in the gene by sequencing nucleic acid molecules from individuals who are thought to harbour a gene variant. Such individuals are people exhibiting the symptoms of a specific disease which is thought to be related to the

dysfunction of a particular gene. Once a gene variant has been sequenced and compared with the wild-type further investigations are then undertaken to examine the cell biology of the protein encoded by the variant gene. The results of these investigations are then examined in the light of the physical symptoms in order to deduce a correlation.

It therefore follows that unravelling the nature of a gene variant and relating it to function and then clinical symptoms is a long and tedious process, especially when one considers that a given gene can be 3.6% polymorphic. It is therefore apparent that simply identifying which variant to investigate further can be a difficult step in itself. This is true not only for the field of human genetics but also in respect of studies of other animal and plant species.

With this in mind, we have developed a novel assay for quickly and efficiently determining the likely significance of a gene variant.

Our novel methodology is based upon the basic structure of proteins.

The basic structural unit of a protein is an amino acid. An amino acid consists of an amino group, a carboxyl group, a hydrogen atom and a distinctive R group bonded to a carbon atom, conventionally known as the side chain. There are 22 amino acids and any number and combination of them are able to join, via peptide bonds, to form a sequence, or chain, of amino acids known as peptides. Thus the sequence of bonds running the length of the peptide chain is known as

the backbone. Additionally, intra and inter peptide chain linkages also exist, for example, in the former instance the amino group of lysine can form a peptide bond with the gamma carboxyl group of glutamic acid; and, in the latter instance, bonds may also exist between side chains of amino acids as a result of the formation of disulphide bonds thus forming crosslinks between separate peptide chains. Adjacent peptide chains can therefore join to form a secondary structure such as dimers or trimers etc. The secondary structures can then fold, due to the nature of the interaction of adjacent amino acids, to form a three dimensional tertiary structure. This tertiary structure represents the active form of the protein and may comprise sites, or pockets, into which other molecules fit in order to activate the protein or allow the protein to respond thereto.

Digestion, or break down, of proteins in a controlled fashion, occurs all the time during the process of alimentary digestion. A class of enzymes known as proteases perform this function. They basically attack specific bonds in order to cleave the protein at sites where these bonds exist. It follows that different proteins will have different susceptibilities to various enzymes depending upon their primary structure.

Whilst all this information is known, no-one has thought to take advantage of it before in relation to genetics and, in particular, in relation to screening a number of genetic variants whose functional, or even clinical, significance is unknown. Accordingly, no-one has thought to use this information as a basis to tackle the large number of genetic variants that exist in order to determine which are the

clinically, or technologically, significant variants.

However, we have used this information to develop a novel assay which can screen any number of variants, simultaneously if required, in order to determine which, if any, require further investigation.

Our methodology is quick, efficient and inexpensive to perform.

Statements of Invention

According to the invention there is therefore provided a method for determining the significance of a given nucleic acid polymorphism or mutation, in a nucleic acid molecule, on the structural properties of a protein encoded by said nucleic acid molecule comprising:

(a) exposing the protein encoded by said nucleic acid molecule to at least one protease; and

(b) determining whether, or to what extent, proteolytic cleavage takes place; and, optionally,

(c) comparing this proteolytic cleavage with that of the wild-type protein when exposed to the same protease(s).

According to a further aspect of the invention there is provided a screening method for determining the significance of a plurality of variants of at least one gene comprising:

(a) obtaining a sample of protein encoded by each of said variants;

(b) exposing each protein to at least one protease;

(c) determining whether, or to what extent, proteolytic cleavage takes place; and

(d) comparing this proteolytic cleavage with that of the wild-type protein when exposed to the same protease(s).

In a preferred embodiment of the invention when the above screening method is employed, the plurality of protein variants are exposed to a plurality of proteases and the corresponding proteolytic cleavage is determined. Most ideally, the screening methodology involves examining the plurality of variants relating to different genes. Thus, in a single batch, the plurality of variants corresponding to multiple genes are examined in respect of at least one protease, and ideally in respect of a plurality of proteases and a determination of proteolytic cleavage is made in respect of the digestion of each variant by each protease.

It will be apparent to those skilled in the art that, using this methodology, where a plurality of proteases are employed a cleavage or digestion profile will be provided in respect of each variant and this parameters ideally can be compared with the digestion profile of the wild-type protein and so used to determine the functional significance of a variant of any one or more of the said genes.

In a preferred embodiment of the invention said protein encoded by said nucleic acid molecule or gene variant is exposed to a plurality of proteases and ideally different proteases which attack different bonds. Proteases that are suitable for

use in the methodology of the invention include: Trypsin, chymotrypsin, proteinase K, aminopeptidase, carboxypeptidase, collagenase, elastase, Kallikrein, metalloendopeptidase, papain, pepsin, and indeed any other known protease.

Notably, where cleavage is different from that exhibited by the wild-type, one would conclude that the variant, or indeed a combination of variants, was significant. This is because the variant(s) would either render the protein more vulnerable to digestion or confer resistance to digestion as a result of alteration(s) to the tertiary, or structural, form of the protein.

In yet a further preferred embodiment of the invention a plurality of proteins encoded by a plurality of genetic variants are tested in parallel and thus the methodology of the invention may be performed as a screening methodology where a plurality of incubation receptacles are filled with a corresponding plurality of proteins to be tested and then said proteins are exposed to a selected protease, or group of proteases, or vice versa, either simultaneously or successively.

More preferably still, the methodology of the invention involves incubating the protein(s) to be tested with the said protease(s) under conditions that support the activity of the relevant enzyme(s). For example, this may involve exposing the test protein to the enzyme at a temperature at which the enzyme is optimally functional, such as 37°C, and for a time sufficient for the enzyme to perform its

activity, for example between 15 minutes and 1.5 hours.

More preferably still, after a suitable length of time the incubation period is terminated, for example, by adding an enzyme inhibitor to the incubation receptacle. Finally, proteolytic cleavage is assessed using any conventional protein assay technique such as, for example, SDS-PAGE analysis either followed by staining the gel (coomassie blue or silver staining) or by western blotting. Optionally, additional studies may then be undertaken to determine the functionality of the protein variant.

In yet a further preferred embodiment of the invention the technique undertaken, in order to determine the extent of proteolytic cleavage, involves assaying not only each test protein but also the wild-type protein that, ideally, has been exposed to the relevant enzyme(s) and, ideally also, a sample of the wild-type and test protein(s) that has not been exposed to the relevant enzyme(s). In this way, a positive and a negative control are included in the assay for the purpose of determining the amount of proteolytic cleavage that the test protein exhibits *vis a vis* the wild-type protein and also the background level of protein degradation experienced as a result of the assay conditions.

It is to be understood that the invention is not to be limited to the specific assay that is chosen to assess proteolytic cleavage, rather the invention, principally, lies in the use of the technique of proteolytic cleavage to assay the likely functional significance of genetic variants.

It follows from the information above regarding the tertiary structure of the protein that the nature of the amino acids in the peptide chain will determine the protein folding and so susceptibility to different enzymes. In turn, the nature of the amino acids in the peptide chain will be determined by the nucleic acid coding sequence and so variations in this sequence will lead to variations at the amino acid level and so differential protein folding and thus variable susceptibility to proteolytic cleavage.

Given that the assay is quick and efficient to perform, a whole range of proteins, each coded by a genetic variant for a given gene, or more than one gene, can be simultaneously assayed in order to determine which variant gives rise to a change in the tertiary structure of the amino acid and thus which is most likely to affect the functioning of the protein.

An embodiment of the invention will now be described by way of example only, with reference to variants in the growth hormone gene (*GH1*) and the following examples.

In Figure 1 there is shown the digestion profile of a number of variants of the growth hormone gene when exposed to trypsin, chymotrypsin or proteinase K.

Experimental Subjects

The experimental subjects in which the mutations for the proteolysis study were identified are those described in the original Human Mutation paper, Millar et al.

Two different patient groups were studied. The first comprised 41 unrelated children of Caucasian origin with short stature (age 1-15 years) who matched the specific selection criteria applied in Cardiff (the "Cardiff criteria") outlined below. Details were taken of family history, clinical and auxological variables, and previously performed laboratory investigations (Table 1). Standard deviation scores (SDS) were calculated for birth weight, height prior to GH treatment, body mass index prior to GH treatment, height velocity immediately prior to GH secretion testing, paternal and maternal heights, and the target for adult height derived from parental measurements (Table 1). The degree of bone age delay and the results of GH secretion tests were also noted. Blood samples for DNA analysis were taken from the index case and appropriate close relatives.

The second group comprised 11 unrelated patients with short stature and idiopathic isolated growth hormone deficiency (IGHD) in whom *GH1* gene deletions had been excluded by Southern blotting. Eight of these individuals came from families with two or more first-degree relatives with IGHD (familial IGHD) whilst 3 individuals represented sporadic cases of IGHD. In only one of the familial IGHD cases (family 37) did short stature appear to segregate as an autosomal dominant trait. Blood samples for DNA analysis were taken from available relatives in each family.

Control DNA samples were obtained from lymphocytes taken from 154 male British army recruits of Caucasian origin who were unselected for height. Height data were available for 124 of these individuals (mean: 1.76 $\pm$ 0.07 m) and the height distribution was found to be normal (Shapiro-Wilk statistic W=0.984, p=0.16). Ethical approval for these studies was obtained from the Multi-Regional Ethics Committee (MREC).

Patient Selection Criteria

The key criterion for inclusion in this study was that the clinician assessing the child should have had sufficient concern with regard to the child's growth pattern to warrant GH secretion testing. The children selected exhibited a clinical phenotype that adhered to the following criteria, henceforth termed "Cardiff criteria":

(a) sufficient clinical concern to have warranted GH secretion testing, *regardless* of the type of test, the test results, or indeed whether the child attended for testing;

(b) no recognisable pathology likely to account for the observed growth failure;

(c) short stature: defined as a predicted height trajectory below the lower limit of an individual's estimated target adult height, based upon the heights of that individual's parents (Tanner and Whitehouse 1976);

(d) height velocity on or below the 25th percentile for age (uncorrected for bone age); and

(e) evidence of bone age delay in those pre-pubertal when compared with chronological age by reference to the Tanner-Whitehouse scale (TW2 method; Tanner et al, 1983). This delay should be of at least two years *except* in children of ≤ 5 years of age.

Materials & Methods

Polymerase chain reaction (PCR) amplification of a GH1-specific fragment

Genomic DNA was extracted from patient lymphocytes by standard procedures. PCR amplification of a 3.2 kb *GH1*-specific fragment was performed as described (2).

Cloning and sequencing of GH1 gene-specific PCR fragments

GH1 gene-specific (3.2 kb) PCR fragments were sequenced directly with BigDye v3.0 (Applied Biosystems, Foster City, CA) and analysed on an ABI 3100 DNA sequencer (Applied Biosystems) as described (2). Additional primers used for sequencing in the reverse direction were GHBFR (5' TGGGTGCCCTCTGGCC 3'; –262 to –278), GHSEQ1R (5' AGATTGGCCAAATACTGG 3'; +215 to +198), GHSEQ2R (5' GGAATAGACTCTGAGAAAC 3'; +785 to +767), GHSEQ3R (5' TCCCTTTCTCATTCATTC 3'; +1281 to +1264), GHSEQ4R (5' CCCGAATAGACCCCGC 3'; +1745 to +1730) [Numbering relative to the transcriptional initiation site at +1; GenBank Accession No. J03071]. Samples containing sequence variants were cloned into pGEM-T (Promega, Madison WI) followed by sequencing of a minimum of four clones per individual.

In vitro expression and assay of biological activity of GH variants

A cloned wild-type *GH1* cDNA incorporating a His tag on the carboxy terminal was modified using site-directed mutagenesis as previously described (3) to generate the GH variants.

This vector was then transfected into High Five insect cells (Invitrogen) as previously described (3), and human GH in the culture supernatants quantified by ELISA (DRG Diagnostics, Marburg, Germany). The cross-reactivity in the ELISA of the GH variants and insect cell-expressed wild-type GH was confirmed by dilutional analysis to be equal to that of the assay reference preparation (calibrated against the MRC 1st IRP 80/505 reference preparation).

Proteolytic digestion of the GH variants

Trypsin, chymotrypsin, or proteinase K (all Sigma, Poole, UK) were added to a final concentration of 0.1μg/ml to 100μl culture medium harvested from insect cells expressing either wild-type GH or a variant (60nM) and then incubated at 37°C for 1 hr. Previous dose-dependent studies on wild-type GH had shown that 0.1μg/ml was the lowest concentration at which GH degradation was detectable by all three enzymes. After the 1 hr treatment period, 10μl trypsin-chymotrypsin inhibitor (500μg/ml) was added to stop the trypsin and chymotrypsin digests and 1μl PMSF (0.1M) was added to stop the proteinase K digest. Each reaction was then incubated for a further 15 mins at 37°C. Samples were analysed by SDS-PAGE on a 12% gel using a mini gel apparatus (Bio-Rad Laboratories, Hercules,

CA). Equivalent amounts of undigested wild-type GH and variant that had been incubated for 1 hour at 37°C were also run on the gel. The gel was electroblotted onto PVDF membrane as previously described (6), probed with a mouse monoclonal anti-human GH antibody (Lab Vision, Fremont, CA), diluted 1:500, detected using an anti-mouse IgG-horse radish peroxidaseHRP (HRP) conjugate (1:5000, Amersham Biosciences) and visualised by enhanced chemiluminescence (ECL Plus, Amersham Biosciences). Films were analysed using the Alpha Imager 1200 digital imaging system (Alpha Innotech Corp, San Leandro, CA) and the results expressed as the amount of GH remaining following enzyme digestion as a percentage of undigested GH. The experiments were repeated 3 times and assessed statistically by a two-tailed t-test.

Molecular modelling

The variants were structurally analysed by inspection of the appropriate variant amino acid residue in the X-ray crystallographic structure of human GH (PDB: 3HHR) [8]. The wild-type and mutant GH structures were compared with respect to electrostatic interactions, hydrogen bonding, hydrophobic interactions and surface exposure. Molecular graphics were performed using the ICM molecular modelling software suite (Molsoft LLC, San Diego, CA).

Results

Proteolysis Studies

Figure 1 shows the results of enzyme analysis performed on a number of GH variants in order to determine which, if any, of these variants alter the structural

properties of the protein and so are likely to interfere with the activity thereof. Twelve variants were examined and it can be seen that with respect to the wild-type (WT), left hand side of the Figure, the majority of these variants have an effect on the susceptibility of the protein to proteolytic digestion. The variants Thr27Ile and Gln91Leu were particularly vulnerable to proteolysis and, in each case, proteolysis proceeded most efficiently using the enzyme chymotrypsin. With reference to Figure 2 it can be seen that the variant Thr27Ile is predicted to affect internal packing around its helix 1 and the loop between helix 2 and helix 3. This obviously has important structural implications which is reflected in the data shown in Figure 1. Similarly, in the case of Gln91Leu the substitution increases hydrophobicity and may affect solubility and folding. This has implications for the structure of the protein and thus its susceptibility to proteolysis.

In contrast, Arg16Cys and Lys41Arg, whilst showing different proteolysis profiles, compared to the wild-type, are less affected than the previous variants. However, for Arg16Cys the predicted structural changes concern inter-molecular bridging rather than adverse effects on the shape of the protein. This could explain why the proteolysis profile is less affected. In the case of Lys41Arg, the variant is thought to conserve ionic interactions but may lead to steric hindrance. Again, the implication here is that the shape of the protein is likely to be conserved.

In contrast, other variants showed only marginal susceptibility to proteolysis

such as VAL110Ile and Thr175Ala which were both most resistant to proteolysis by the enzyme chymotrypsin.

The results of this study show that GH variants can be characterised in terms of their proteolysis signature in response to selected proteases and this information represents a first step towards selecting clinically and technologically important variants for further analysis.

Table 1

Children with short stature adhering to the specific ("Cardiff") selection criteria; clinical and auxological variables, and laboratory investigations.

P	S	BW	GH	T	H	HA	BA	BMI	HV	Mat	Pat
10	**M**	**-2.4**	**38.7**	**E**	**-1.8**	**8.9**	**7.0**	**-1.4**	**-0.9**	**0.9**	**-2.6**
12	**F**	**-2.7**	**X**	**X**	**-1.8**	**1.0**	**0.5**	**-3.6**	**-4.3**	**-1.5**	**-0.5**
20	**M**	**2.1**	**4.1**	**I**	**-3.4**	**7.3**	**5.3**	**1.0**	**-1.8**	**-1.8**	**0.1**
33*	**M**	**-0.4**	**X**	**X**	**-2.1**	**6.5**	**3.9**	**-0.7**	**-1.5**	**-1.0**	**-0.2**
53	**M**	**-3.2**	**27.2**	**C**	**-3.8**	**13.5**	**11.5**	**-4.9**	**-1.3**	**-2.8**	**-1.5**
57	**F**	**-3.4**	**27.3**	**R**	**-2.8**	**2.7**	**1.5**	**-1.7**	**-1.7**	**0.3**	**0.7**
62	**F**	**-1.9**	**27.0**	**G**	**-4.4**	**4.6**	**3.5**	**0.6**	**-0.7**	**-1.5**	**-1.9**
63	**M**	**-0.9**	**1.3**	**G**	**-4.5**	**8.2**	**6.2**	**-0.2**	**-4.4**	**0.6**	**-1.1**
66	**M**	**-1.0**	**18.8**	**I**	**-2.9**	**8.1**	**6.1**	**-0.3**	**-1.2**	**-3.9**	**-1.6**
71§	**F**	**-2.9**	**4.0**	**I**	**-1.4**	**3.7**	**3.7**	**-0.7**	**-0.5**	**-2.4**	**-2.4**
75	**F**	**0.2**	**6.8**	**C**	**-6.2**	**14.1**	**13.5**	**4.2**	**-3.5**	**-0.6**	**0.3**
76§	**F**	**-1.8**	**18.3**	**I**	**-2.2**	**8.0**	**7.0**	**-1.3**	**-1.8**	**-2.4**	**-2.2**
79	**M**	**-3.6**	**28.8**	**C**	**-5.3**	**3.3**	**1.0**	**-1.2**	**-1.1**	**-1.9**	**-1.1**
83	**M**	**-2.8**	**N**	**R**	**-3.5**	**2.0**	**1.0**	**-1.3**	**-2.1**	**2.1**	**0.7**
1	F	1.4	48.6	I	-3.9	13.1	11.1	-0.9	0.1	-3.2	-1.6
2	F	-1.4	20.2	I	-3.0	15.1	13.1	1.1	-1.0	-0.1	-1.7
3	F	-1.2	3.7	C	-3.9	2.8	1.0	2.3	-2.2	-0.8	-1.1
4	M	0.3	26.7	I	-3.4	7.5	4.7	1.4	-2.2	-1.7	-1.2
5	M	-1.7	7.7	I	-3.0	9.0	6.0	1.1	-0.8	-1.5	-0.2
6	M	-3.1	28.4	C	-4.2	5.8	3.0	-0.9	-1.5	-1.9	-3.7
7	F	-2.7	111.3	C	-3.0	5.6	2.5	-2.8	0.4	0.8	-1.2
9	F	0.6	9.8	I	-2.0	13.3	11.3	2.6	1.6	-0.8	-1.0
11	M	0.3	0	C	-4.8	5.8	3.4	-0.1	-1.7	-4.1	-1.4
13	F	-3.1	16.0	C	-4.7	4.3	2.5	-1.4	-4.0	0.2	-1.3
14	M	-0.5	13.2	C	-1.9	12.8	8.8	-1.2	-0.7	-0.2	-1.0
15	F	-0.9	0	C	-5.0	5.0	2.0	0.8	-5.0	-0.2	-0.8
16	M	-0.5	4.6	I	-3.1	6.8	4.8	-1.0	-2.1	-0.3	-0.6
18	M	-3.8	18.0	I	-4.9	8.8	4.8	-1.2	-2.3	-1.1	-0.8
19	F	-0.5	2.1	I	-2.2	4.3	2.0	-0.1	-3.0	-0.5	0.2
26	M	-0.2	38.6	C	-2.8	12.0	9.0	-2.1	-0.1	-0.2	-1.0
27	F	-1.4	5.6	I	-3.9	13.4	10.5	0.2	0.2	-2.1	-1.8
34	F	-5.4	10.0	R	-3.8	1.4	1.0	-2.6	-0.8	-0.1	-0.0
40	M	-1.4	X	X	-1.4	12.8	10.2	-0.4	-1.2	1.0	-0.8
58	M	-3.8	<1	I	-4.5	8.8	6.8	-2.4	-6.6	-3.8	-1.6
59	F	-0.1	<1	I	-3.6	8.0	5.7	-2.6	-0.9	-1.5	-0.8
68	M	-0.9	X	X	-3.1	8.0	5.0	-0.1	-1.2	-1.1	-1.2
77	M	-0.6	48.0	C	-3.9	5.7	2.8	-0.8	-1.6	-0.3	-1.3
80	M	-3.6	33.9	C	-4.2	3.5	1.3	-3.0	-0.8	-0.6	-0.5
81	M	-1.8	N	R	-3.4	5.0	2.3	-3.1	-2.3	-0.1	-0.5
82	F	-1.9	24.2	C	-6.8	2.0	1.0	0.3	-2.6	-0.3	0.4
84	F	-0.3	45.0	C	-4.0	5.8	3.5	-0.1	-3.2	-1.5	0.8

Key. P: patient number; S: sex; BW: birth weight standard deviation score (SDS); GH: GH secretion test result (mIU/L); N: "normal", T: GH secretion test type: I: insulin tolerance test, C: clonidine, G: glucagon, E: exercise, R: random, X: test declined. H: height SDS; HA: age at height SDS and bone age assessment; BA: bone age in years; BMI: body mass index SDS; HV: height velocity SDS; Mat: maternal height SDS; Pat: paternal height SDS.

Data from patients possessing *GH1* gene lesions are shown in bold type.

*one of two siblings with a similar phenotype.

§family history of GHD.

References

1 Ranke MB 1996 Towards a consensus on the definition of idiopathic short stature. Horm Res 45 Suppl. 2:64-66

2 Horan M, Millar DS, Hedderich J, Lewis G, Newsway V, Mo N, Fryklund L, Procter AM, Krawczak M, Cooper DN 2003 Human growth hormone 1 (*GH1*) gene expression: complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. Hum Mutat 21:408-423

3 Millar DS, Lewis MD, Horan M, Newsway V, Easter TE, Gregory JW, Fryklund L, Norin M, Crowne EC, Davies S, Edwards P, Kirk J, Waldran K, Smith PJ, Phillips III JA, Scanlon MF, Krawczak M, Cooper DN, Procter AM 2003 Novel mutations of the growth hormone 1 (*GH1*) gene disclosed by modulation of the clinical selection criteria for individuals with short stature. Hum Mutat 21:424-440

4 Ross RJ, Esposito N, Shen XY, von Laue S, Chew SL, Dobson PR, Postel-Vinay MC, Finidori J 1997 A short isoform of the human growth hormone receptor functions as a dominant negative inhibitor of the full-length receptor and generates large amounts of binding protein. Mol Endocrinol 11:265-273

5 von Laue S, Finidori J, Maamra M, Shen X-Y, Justice S, Dobson PRM, Ross RJM 2000 Stimulation of endogenous GH and interleukin-6 receptors

selectively activates different Jaks and Stats, with a Stat5-specific synergistic effect of dexamethasone. J Endocrinol 165:301-311

6 Lewis MD, Ham J, Rees DA, Lewis BM Scanlon MF 2002 Mitogen-activated protein kinase mediates epidermal growth factor-induced morphogenesis in pituitary GH_3 cells. J Neuroendocrinol 14:361-367

7 Green H, Kehinde O 1976 Spontaneous heritable changes leading to increased adipose conversion in 3T3 cells. Cell 7:105-113

8 de Vos AM, Ultsch M, Kossiakoff AA 1992 Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. Science 255:306-312

9 Krawczak M, Chuzhanova NA, Cooper DN 1999 Evolution of the proximal promoter region of the mammalian growth hormone gene. Gene 237:143-151

10 Clackson T, Wells JA 1995 A hot spot of binding energy in the hormone-receptor interface. Science 267:383-386

11 Clackson T, Ultsch MH, Wells JA, de Vos AM 1998 Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. J Mol Biol 277:1111-1128

12 Cunningham BC, Wells JA 1989 High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 244:1081-1084

13 Hansen JH, Wang X, Kopchick JJ, Bouchelouche P, Neilson JH, Galsgaard ED, Billestrup N 1996 Identification of tyrosine residues in the intracellular domain of the growth hormone receptor required for transcriptional signalling and STAT 5 activation. J Biol Chem 271:12669-12673

14 Sotiropoulos A, Perrot-Applanat M, Dinerstein H, Pallier A, Postel-Vinay M-C, Finidori J, Kelly PA 1994 Distinct cytoplasmic regions of the growth hormone receptor are required for activation of JAK2, mitogen-activated protein kinase, and transcription. Endocrinology 135:1292-1298

15 15 VanderKuur J, Allevato G, Billestrup N, Norstedt G, Carter-Su C 1995 Growth hormonr-promoted tyrosyl phosphorylation of SHC proteins and SHC association with Grb2. J Biol Chem 270:7587-7593

16 VanderKuur JA, Butch ER, Waters SB, Pessin JE, Guan KL, Carter-Su 1997 Signalling molecules involved in coupling growth hormone receptor to mitogen-activated protein kinase activation. Endocrinology 138:4301-4307

17 Liang L, Jiang J, Frank SJ 2000 Insulin receptor substrate-1 enhancement of growth hormonr-induced MAP kinase activation. Endocrinology 141:3328-3336

18 Kim S-O, Loesch K, Wang X, Jiang J, Mei L, Cunnick JM, Wu J, Frank SJ 2002 A role for Grb2-associated binder-1 in growth hormone signalling. Endocrinology 143:4856-4867

19 Yamauchi T, Ueki K, Tobe K, Tamemoto H, Sekine N, Wada M, Honjo M, Takahashi M, Takahashi T, Hirai H, Tushima T, Akanuma Y, Fujita T, Komuro I, Yazaki Y, Kadowaki T 1997 Tyrosine phosphorylation of the EGF receptor by the kinase Jak2 is induced by growth hormone. Nature 390:91-96

20 Zhu T, Ling L, Lobie PE 2002 Identification of a JAK2-independent pathway regulating growth hormone (GH)-stimulated p44/42 mitogen-activated protein kinase activity. J Biol Chem 277:46692-45603

21 Shoba LNN, Newman M, Liu W, Lowe WL 2001 LY 294002, an inhibitor of phosphatidylinositol 3-kinase, inhibits GH-mediated expression of the IGF-1 gene in rat hepatocytes. Endocrinology 142:3980-3986

22 Frago LM, Pañeda C, Dickson SL, Hewson AK, Argente J, Chowen JA 2002 Growth hormone (GH) and GH-releasing peptide-6 increase insulin-like

growth factor-I expression and activate intracellular signalling pathways involved in neuroprotection. Endocrinology 143:4113-4122

23 Metherell LA,, Akker SA, Munroe PB, Rose SJ, Caulfield M, Savage MO, Chew SL, Clark AJL 2001 Pseudoexon activation as a novel mechanism for disease resulting in atypical growth-hormone insensitivity. Am J Hum Genet 69:641-646

24 Milward AD, Maamra M, Wilkinson IR, Metherell LA, Savage MO, Clark AJL, Ross RJM 2003 A mutated growth hormone receptor (GHR1-656) with an extracellular domain insertion activates STAT 5 but not MAP kinase. Program of the 85th Annual Meeting of The Endocrine Society, Philadelphia, PA, 2003, p62 (Abstract)

CLAIMS

1. A method for determining the significance of a given nucleic acid polymorphism or mutation, in a nucleic acid molecule, on the structural properties of a protein encoded by said nucleic acid molecule comprising:

(a) exposing the protein encoded by said nucleic acid molecule to at least one protease; and

(b) determining whether, or to what extent, proteolytic cleavage takes place; and, optionally,

(c) comparing this proteolytic cleavage with that of the wild-type protein when exposed to the same protease(s).

2. A screening method for determining the significance of a plurality of variants of at least one gene comprising:

(a) obtaining a sample of protein encoded by each of said variants;

(b) exposing each protein to at least one protease;

(c) determining whether, or to what extent, proteolytic cleavage takes place; and

(d) comparing the proteolytic cleavage with that of the wild-type protein when exposed to the same protease(s).

3. A method according to claim 1 or 2 wherein said protein is exposed to a plurality of proteases.

4. A method according to claim 3 wherein at least some of said proteases and so attack different sites within the protein.

5. A method according to any preceding claim wherein said protease(s) comprises any one or more of the following: trypsin, chymotrypsin, proteinase K, aminopeptidase, carboxypeptidase, collagenase, elastase, Kallikrein, metalloendopeptidase, papain or pepsin.

6. A method according to any preceding claim wherein a plurality of proteins are exposed to said protease(s).

7. A method according to claims 3-6 wherein said proteins are exposed to said proteases, or vice versa, simultaneously.

8. A method according to any preceding claim wherein said protein(s) is exposed to said different proteases either simultaneously or successively.

9. A method according to any preceding claim wherein said protein(s) are exposed to said protease(s) under conditions that support the activity of said protease(s).

10. A method according to any preceding claim wherein digestion of said protein(s) is terminated by adding at least one protease inhibitor to the reaction.

11. A method according to any preceding claim wherein proteolytic cleavage is determined using a conventional protein assay.

12. A method according to claim 11 wherein said assay involves SDS-PAGE analysis.

13. A method according to claim 12 wherein said analysis is followed by staining or blotting.

14. A method according to any preceding claim wherein additional studies are undertaken to determine the functionality of the protein variant.

15. A method according to any preceding claim wherein part (a) involves further exposing the wild-type protein to said at least one protease and part (b) involves determining whether and to what extent proteolytic cleavage of said wild-type protein takes place.

16. A method according to any preceding claim wherein the wild-type protein and, optionally, the variant protein are subjected to the conditions of the proteolytic reaction, in the absence of the said protease(s), and then the extent of proteolytic cleavage is determined.

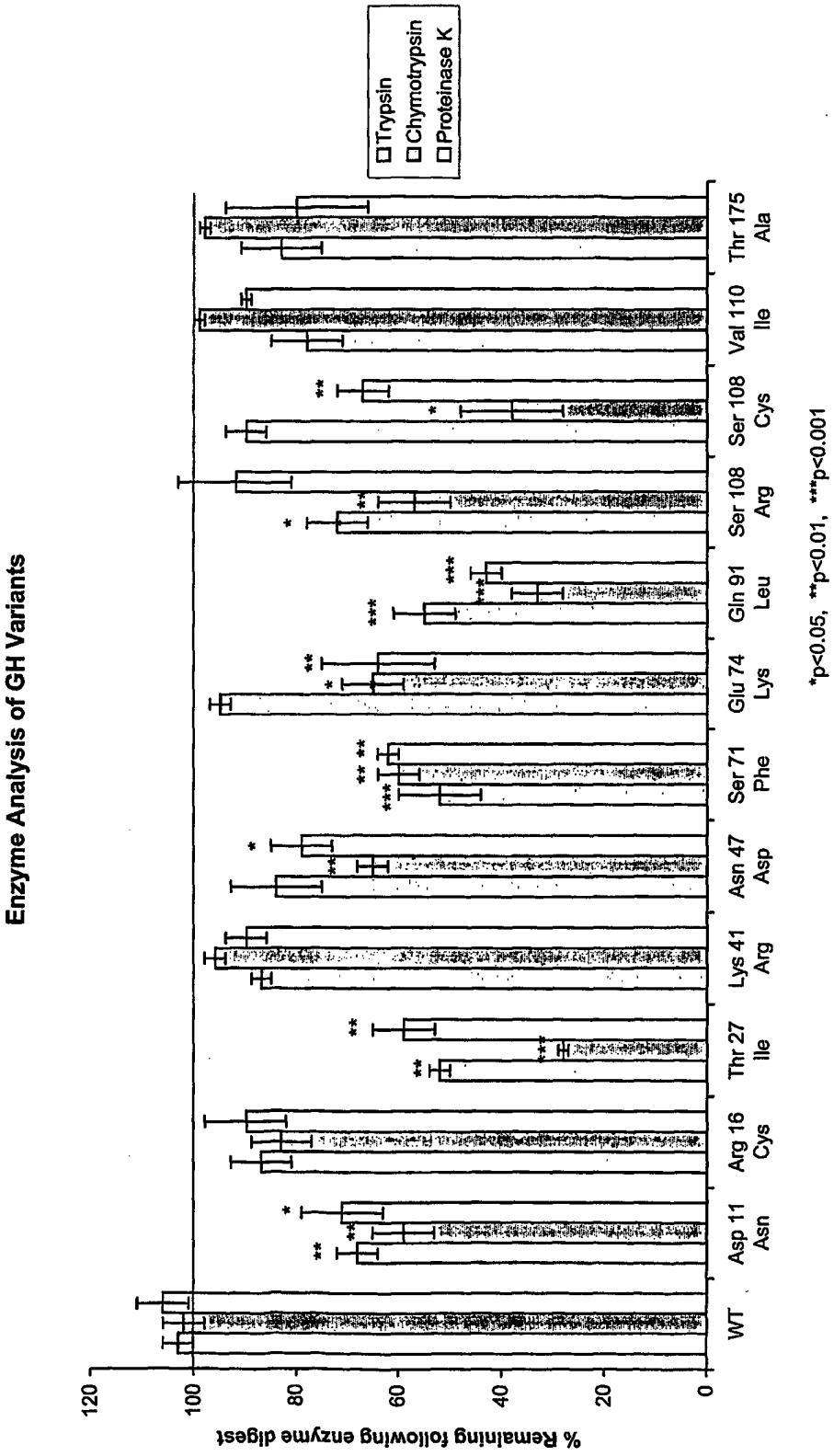
Enzyme Analysis of GH Variants
Trypsin
Chymotrypsin
Proteinase K
% Remaining following enzyme digest
120
100
80
60
40
20
0
WT
Asp 11 Asn
Arg 16 Cys
Thr 27 Ile
Lys 41 Arg
Asn 47 Asp
Ser 71 Phe
Glu 74 Lys
Gln 91 Leu
Ser 108 Arg
Ser 108 Cys
Val 110 Ile
Thr 175 Ala
*p<0.05, **p<0.01, ***p<0.001

FIGURE 1

LOCATION OF AMINO ACID RESIDUES INVOLVED IN MISSENSE MUTATIONS AND THE PREDICTED CONSEQUENCES OF SUBSTITUTION

Amino Acid Substitution	Location of amino acid residues, interactions and consequences of substitution
D11N	D11 (helix 1) is solvent accessible. No obvious adverse consequences of substitution by N.
R16C	R16 (helix 1) is solvent accessible and interacts with site 2 andE44 & W169 of GHR. Substitution by C could adversely affect site 2 binding. Introduction of unpaired cysteine could also lead to inter-molecular disulphide bridging with unpaired cysteine could also lead to inter-molecular disulphide bridging with consequent protein aggregation.
T27I	T27 lies in helix 1 but is buried. Substitution by I may affect internal packing around helix 1 and the loop between helix 2 and helix 3.
K41R	K41 (loop 1 lies between helices 1 and 2) is solvent accessible. K41 N atom exhibits ionic interaction with GHR E127 Oε2 and has been implicated in GHR binding by alanine scanning mutagenesis. Substitution by R may conserve ionic interaction but could cause unfavourable steric interactions.
N47D	N47 (loop 1) is solvent accessible. No interaction with GHR. No obvious adverse consequences of substitution by D.
S71F	S71 (loop 1) is solvent accessible. Hydrophobic side-chain of substituting F could decrease protein solubility and affect folding.
E74K	E74, partially exposed with N-terminal of helix 2, may interact with Q137 thereby stabilising helix 2. Introduction of K may affect helix stability.
Q91L	Q91 located at C-terminal end of helix 2. Introduction of L increases hydrophobicity and may affect solubility and folding.
S108R	S108 (helix 3) is solvent accessible but does not interact with GHR. Substitution by R could adversely affect helix formation.
S108C	S108 (helix 3) is solvent accessible. Introduction of unpaired cysteine could lead to inter-molecular disulphide bonding with consequent protein aggregation.
V110I	V110 (N-terminal of helix 3) is deeply buried in hydrophobic core. Conservative substitution but I has longer sidechain and may encounter steric hindrance.
T175A	T157 (helix 4) is solvent accessible and has been implicated in GHR binding by alanine scanning mutagenesis. T175 forms H-bond with D171 of GH and W169 and R43 of GHR. Introduction of A may destabilise helix thereby decreasing receptor binding.

FIGURE 2

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB 03/04740

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data, EPO-Internal, MEDLINE, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GAUCZYNSKI S ET AL: "Recombinant human prion protein mutants huPrP D178N/M129 (FFI) and huPrP +90R (fCGD) reveal proteinase K resistance" JOURNAL OF CELL SCIENCE, vol. 115, no. 21, 1 November 2002 (2002-11-01), pages 4025-4036, XP002271831 page 4026 -page 4029; figure FIG.2 page 4034, right-hand column	1,2,5-16

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 February 2004	11/03/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Stoyanov, B

Form PCT/ISA/210 (second sheet) (July 1992)